



## King's Research Portal

DOI:

[10.1021/ja4085034](https://doi.org/10.1021/ja4085034)

*Document Version*

Peer reviewed version

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Cornell, T. A., Fu, J., Newland, S. H., & Orner, B. P. (2013). Detection of Specific Protein-Protein Interactions in Nanocages by Engineering Bipartite FIAsh Binding Sites. *Journal of the American Chemical Society*, 135(44), 16618-16624. <https://doi.org/10.1021/ja4085034>

### Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### Take down policy

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

Article

## Detection of Specific Protein-Protein Interactions in Nanocages by Engineering Bipartite FIAsh Binding Sites

Thomas A. Cornell, Jing Fu, Stephanie H. Newland, and Brendan Patrick Orner

*J. Am. Chem. Soc.*, **Just Accepted Manuscript** • DOI: 10.1021/ja4085034 • Publication Date (Web): 08 Oct 2013

Downloaded from <http://pubs.acs.org> on October 13, 2013

### Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.



ACS Publications  
High quality. High impact.

Journal of the American Chemical Society is published by the American Chemical Society, 1155 Sixteenth Street N.W., Washington, DC 20036  
Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

# Detection of Specific Protein-Protein Interactions in Nanocages by Engineering Bipartite FIAsh Binding Sites

Thomas A. Cornell<sup>†, Δ</sup>, Jing Fu<sup>Δ, ‡</sup>, Stephanie H. Newland<sup>#, ‡</sup>, and Brendan P. Orner<sup>†, \*</sup>

<sup>†</sup> Department of Chemistry, King's College London, London, SE1 1DB, United Kingdom. <sup>Δ</sup> Division of Chemistry and Biological Chemistry, Nanyang Technological University, Singapore, 637371. <sup>#</sup> School of Chemistry, University of Southampton, Southampton, SO17 1BJ, United Kingdom.

**KEYWORDS** *Ferritin, DNA binding protein from starved cells, Protein cage, FIAsh, Bipartite tetracysteine display, Self-assembly, Protein engineering, Protein-protein interactions.*

**ABSTRACT:** Proteins that form cage-like structures have been of much recent cross-disciplinary interest due to their application to bioconjugate and materials chemistry, their biological functions spanning multiple essential cellular processes, and their complex structure, often defined by highly symmetric protein-protein interactions. Thus, establishing the fundamentals of their formation, through detecting and quantifying important protein-protein interactions, could be crucial to understanding essential cellular machinery, and for further development of protein-based technologies. Herein we describe a method to monitor the assembly of protein cages by detecting specific, oligomerization state dependent, protein-protein interactions. Our strategy relies on engineering protein monomers to include cysteine pairs that are presented proximally if the cage state assembles. These assembled pairs of cysteines act as binding sites for the fluorescent reagent FIAsh which, once bound, provides a readout for successful oligomerization. As a proof of principle, we applied this technique to the iron storage protein, DNA-binding protein from starved cells (Dps) from *E.coli*. Several linker lengths and conformations for the presentation of the cysteine pairs were screened to optimize the engineered binding sites. We confirmed that our designs were successful in both lysates and with purified proteins, and that FIAsh binding was dependent upon cage assembly. Following successful characterization of the assay, its throughput was expanded. A two-dimension matrix of pH and denaturing buffer conditions was screened to optimize nanocage stability. We intend to use this method for the high throughput screening of protein cage libraries and of conditions for the generation of inorganic nanoparticles within the cavity of these and other cage proteins.

## Introduction

Large, hollow, and often symmetric, cage-like protein assemblies, like ferritin nanocages and virus capsids, provide impetus for investigations into protein folding, protein-protein interactions and self-assembly, all of which underpin protein quaternary structure.<sup>1,2,3</sup> On a biofunctional level, similar protein nanostructures are involved in sequestering metals,<sup>4</sup> creating size-specific pockets of a hydrophobic environment to assist in protein folding,<sup>3</sup> catalyzing the generation of metabolites,<sup>5</sup> as well as delivering and protecting viral genomes.<sup>6</sup> Along with having fundamental importance, protein cages have been the focus of much applied research. To date, protein cages have been used as size constraining reaction vessels for the construction of inorganic materials, and for several potential bio-medically relevant applications such as drug and siRNA delivery.<sup>7, 8, 9, 10, 11, 12, 13</sup>

Many high resolution protein cage structures are available.<sup>14, 15</sup> These structures have paved the way for rational engineering and design, a pursuit that is important for enhancing the properties of protein cages to match those required for further applications. One successful design strategy, applied in different ways, has been to enhance the protein-protein interactions between monomers.<sup>16, 17, 18, 19</sup> However, this type

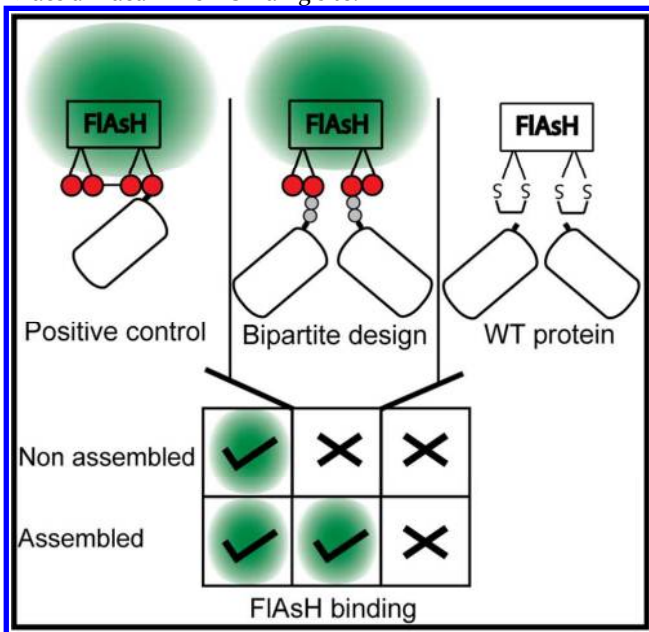
of research can be protracted due to the necessity of iteratively purifying each mutant followed by extensive biophysical characterization with techniques that are often not directly related to cage formation.<sup>2, 20</sup>

Ferritins are ubiquitous protein cages whose structure has been extensively studied due to its relatively straightforward folding and assembly. They store cellular iron through mineralization inside their hollow cavity. *E.coli* DNA-binding protein from starved cells (Dps), a mini-ferritin, has a 9 nm outer diameter and assembles from twelve identical monomers.<sup>21, 22, 23, 24, 25, 26</sup> Each monomer folds into a four helix bundle with an additional helix along the loop between the second and third helix of the bundle (the "BC helix").<sup>17, 27</sup> During ferritin self-assembly, monomers rarely persist and for most ferritins, a 2-fold symmetric dimer is believed to be the most prevalent intermediate. This dimer is thought to be the fundamental building block for cage formation.<sup>21</sup> In Dps, these intermediates are most likely anti-parallel dimers. A consequence of this is that the termini of each monomer projects away from those of the other monomer; only with increased oligomerization to the fully formed cage state do the termini converge. (Figure 2A) Although ferritins have been pursued extensively for applications in material sci-

ence,<sup>10</sup> only minimal work has been performed to optimize the properties of the ferritins for these applications.

Unfortunately, there are few methods for determining the oligomerization states of the cages other than expressing, purifying and assaying each protein followed by mostly low throughput biophysical techniques, such as size exclusion chromatography (SEC), dynamic light scattering (DLS) and transmission electron microscopy (TEM).<sup>2, 16-17, 20, 28</sup> Creating a system that can rapidly identify specific oligomerization states *in vivo* or in cell lysates, would greatly advance research on protein nanocages and protein self-assembly in general.

An ideal method would be one that could distinguish specific protein-protein interactions during the assembly process. A direct oligomerization assay employing a biarsinical fluorescent reagent,<sup>29, 30</sup> of which FIAsh and ReAsH are the most common, is one possibility. These reagents, which have been used as an alternative to GFP variants for protein labeling,<sup>31, 32, 33, 34</sup> exploit the affinity of their arsenic atoms for sulfur atoms in a protein. The reagents bind selectively to proteins with four appropriately presented cysteines (SI Figure S1). This binding results in fluorescence most likely due to a change in rotational properties about the carbon-arsenic bond.<sup>35, 36</sup> Originally, it was suggested that FIAsh ideally interacts with cysteines displayed on a single face of an alpha helix, but later it was shown that the sequence CCPGCC provides an ideal FIAsh binding site.



**Figure 1** – Fluorescence (green) from the reagent, FIAsh, is assembly state dependent only in the bipartite design because of the need to have two sets of cysteine pairs (red) to form a binding site. The positive control provides a binding site that is neither folding, nor oligomerization dependent whereas the negative control provides no FIAsh binding. This is the basis of our strategy to directly monitor protein cage formation.

Recently, it has been shown that the four cysteines in the FIAsh binding motif can be split into two ‘bipartite’ cysteine

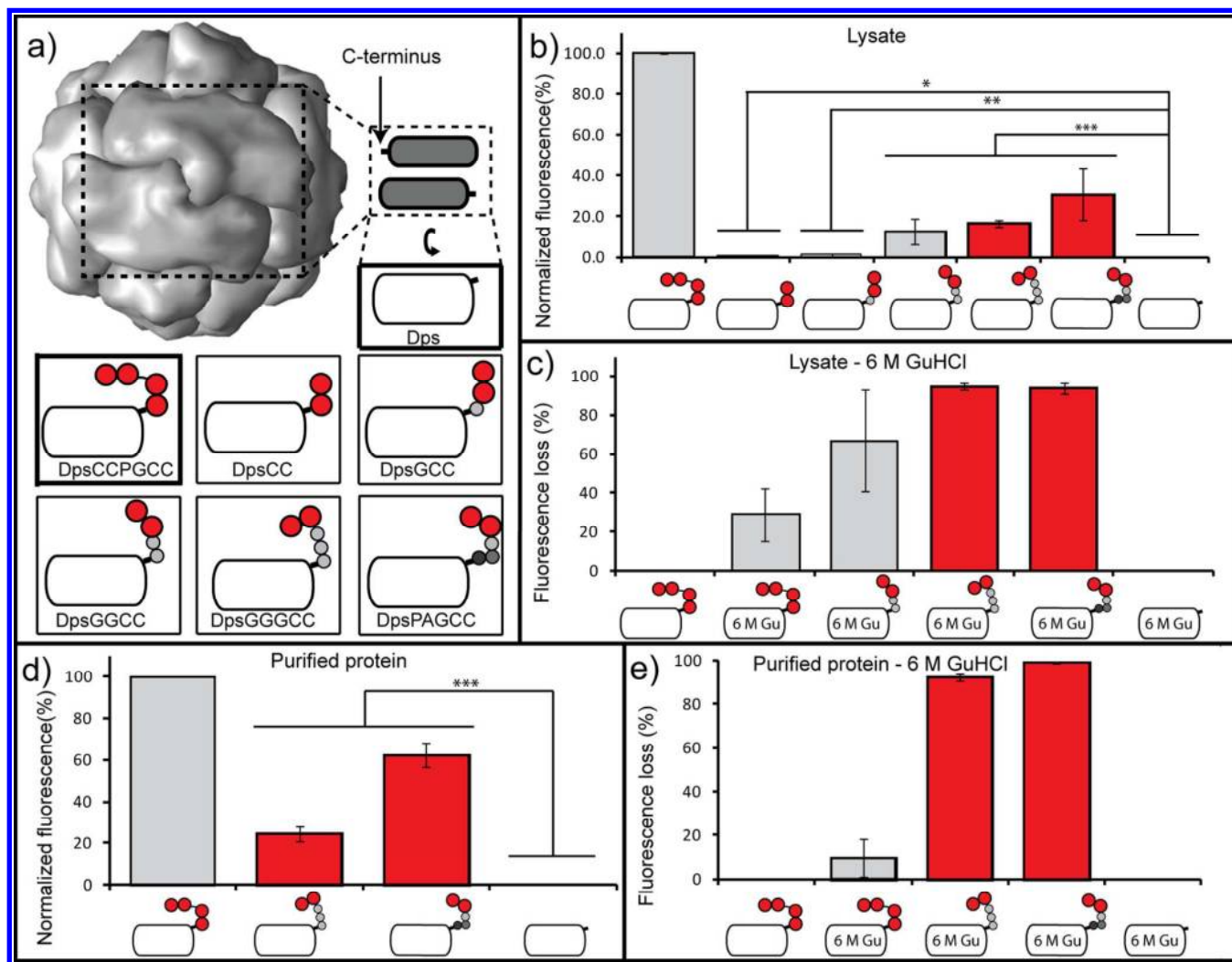
pairs. If the two pairs are positioned apart from each other in the polypeptide sequence, FIAsh can be used to detect when they become proximal during protein folding. Similarly, the two pairs can be placed on two separate polypeptides, and FIAsh can be used to monitor the formation of a protein-protein interaction if the cysteines are positioned appropriately across the interface.<sup>35, 37, 38</sup> Recently this strategy was used to elucidate conformationally transduced signals through the cellular membrane and to provide an explanation for the divergent signaling outcomes of an EGF receptor that dimerizes through coiled-coil motifs.<sup>39</sup>

The first step in developing the biarsinical reagents into probes for protein cage assembly is to design binding sites that only appear upon cage formation. This goal is made more challenging by the fact that the monomers of protein cages, such as Dps, use multiple interfaces for cage formation and various states could be intermediates along the oligomerization pathway. Thus, a design with ultimate utility would be robust enough to distinguish between these oligomerization states. One possible solution therefore would be to design ferritin binding sites that exploit the divergence of monomer termini in the antiparallel dimer intermediate (see above). In the cage form, these termini converge. Thus, terminally appended cysteines would provide a sulfur rich area for a FIAsh binding site that only forms upon cage assembly and not in a dimer intermediate. One must also consider the location of native cysteine residues. Dps has a single native cysteine that is 23.33 Å from the corresponding cysteine in the nearest monomer when assembled into a cage (SI Figure S2). While this will not impede this investigation, consideration must be taken when applying this approach to other proteins.

In our design, we placed cysteine pairs at the Dps C-terminus. (Figure 1) A series of mutants were generated to optimize the binding site; these mutants differ in how the two cysteines are displayed by the monomer either based on linker length (DpsCC, DpsGCC, DpsGGCC, DpsGGGCC) or conformation (DpsPAGCC). (Figure 2a) Extending the C-terminus to display the peptide sequence CCPGCC which includes all four cysteines of an ideal binding site on a single monomer (see above) resulted in the positive control (DpsCCPGCC) which would require neither folding nor assembly to generate a FIAsh signal. We used wild type Dps, which has no additional cysteines, as a negative control (Dps).

## Results and Discussion

An oligomerization state assay with the most utility would be able to evaluate multiple samples rapidly. Therefore, instead of initially working with purified proteins, we used cellular lysates. This decision increased our throughput, allowing the screening of more potential designs. In addition, the use of complex solutions allowed us to screen for selectivity and the success of this method in lysates would demonstrate specificity in the presence of metal ions and sulfur containing metabolites and proteins. We employed conditions containing 1, 2 ethanedithiol (EDT), 2-mercaptoethanol (2-ME), tris(2-carboxyethyl)phosphine (TCEP) and ethylenediaminetetraacetic acid (EDTA) to control oxidation state and metal ion concentration consistent



**Figure 2** – (a) Dps crystal structure (PDB: 1DPS) emphasizing a dimer subunit and divergent presentation of the C-termini along with schematics representing proteins used in this study. The negative control (**Dps**) has no appended C-terminal cysteines whereas the monomer of the positive control (**DpsCCPGCC**) displays a full binding site. The other proteins present a pair of cysteines with a variety of flexible and constrained linkers. Cysteine (red), glycine (light grey), alanine (dark grey), proline (black) (b) FIAsh fluorescence, normalized to controls, for the proteins overexpressed as cell lysates diluted to the same total protein concentration (0.1 mg/ml). (c) Percent fluorescence loss for the proteins in lysates in denaturing conditions. (d) FIAsh fluorescence, normalized to controls, for the proteins overexpressed and purified, diluted to the same protein concentration (0.1 mg/ml) (e) Percent fluorescence loss for the purified proteins in denaturing conditions. The lysate and purified protein data are from six and eight replicates respectively. Error bars are S.D. \* Two-tailed P-values = 0.4804. \*\* Two-tailed P-values = 0.020050. \*\*\* Two-tailed P-values = 0.0001.

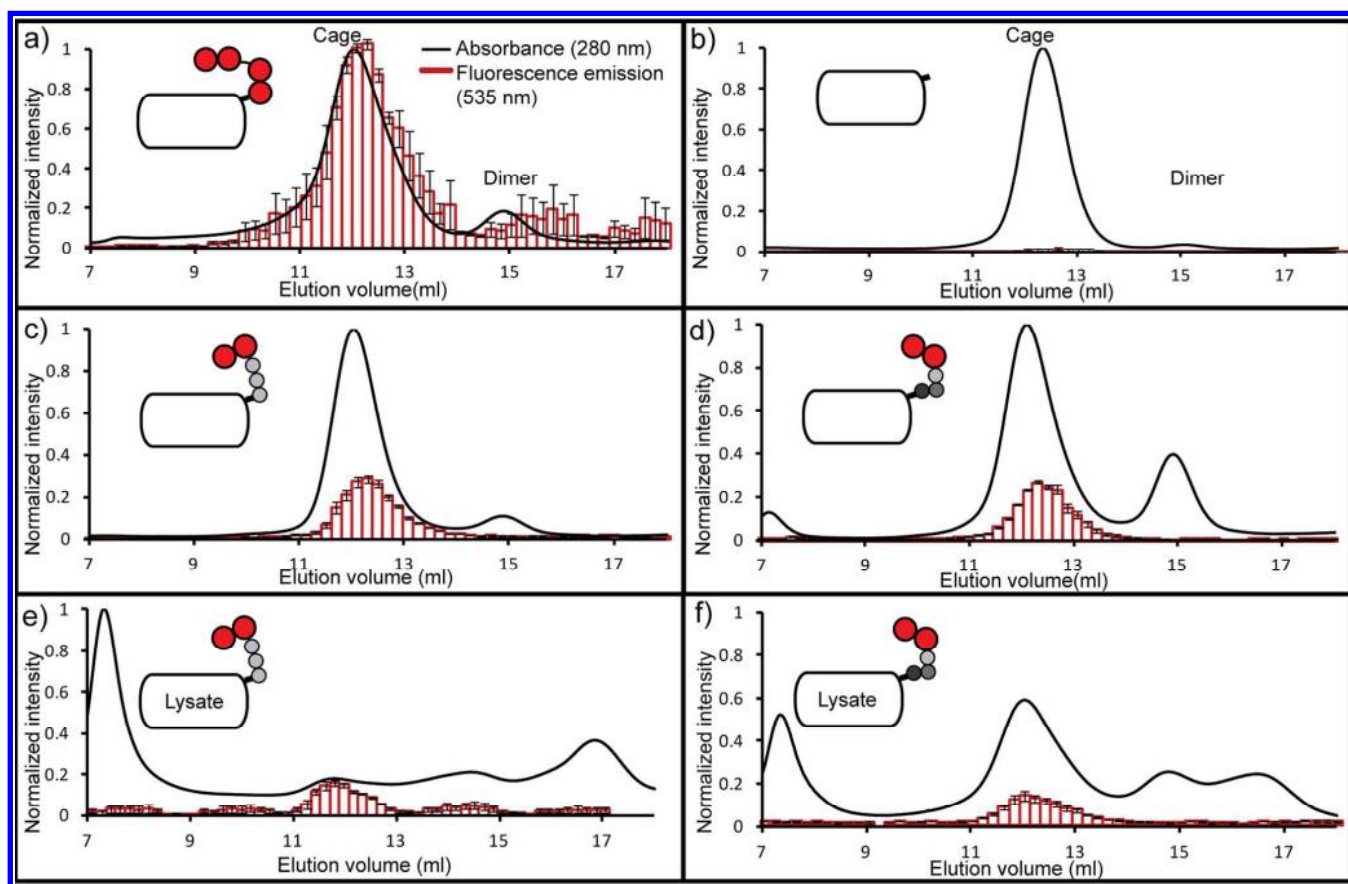
with the literature,<sup>29, 30, 37, 38</sup> however, our conditions are not fully optimized. In lysates, (Figure 2b) the positive control, **DpsCCPGCC**, generated robust fluorescence with added FIAsh, and the negative control (**Dps**) gave nearly undetectable signal. Of the bipartite designs, **DpsCC** and **DpsGCC** had a very weak response whereas **DpsGGCC**, **DpsGGGCC**, and **DpsPAGCC** gave fluorescence that was between 15 and 35% of the positive control suggesting that a longer linker is ideal and one that is more rigid may be optimal.

To confirm that the FIAsh binding is indeed dependent on cage formation, we subjected the lysates to similar experi-

ments but in denaturing conditions. (Figure 2c) As expected the negative control, **Dps**, exhibited no change in fluorescence upon denaturation.

The positive control, **DpsCCPGCC**, was also expected to have no change, however, it did generate a small, but significant loss which may be attributed to restricted access to the tetracysteine tag in the cage and may be suggestive that more optimized equilibration times could enhance the assay. As expected, the





**Figure 3** – Size exclusion chromatograms monitored at absorbance of 280 nm (protein) and fluorescence at 535 nm (bound FAsH) to determine if the assay is selective for the cage oligomerization state. (a-d) Purified proteins (0.5 ml injection of 1 mg/ml). (e, f) Lysate solutions with overexpressed designed protein (0.5 ml injection of 1 mg/ml total protein). All absorbance and fluorescence signals are normalized to the positive control. Results for both the UV and fluorescence traces are averages of three runs.

bipartite designs displayed a large loss in fluorescence upon denaturation with **DpsGGGCC** and **DpsPAGCC** exhibiting nearly a 100% loss in signal with added denaturant. To verify these results, the leading proteins were expressed, purified (SI Figure S1 for sequencing data) and fully characterized (SI Figure S4-8 and SI Table S4 for gels and mass spec, SEC, TEM, CD) and were subjected to the binding experiment and subsequent denaturation. (Figures 2d and 2e) Again, the controls behaved as expected with the exception of the positive control that, again, lost a small amount of signal upon denaturation.

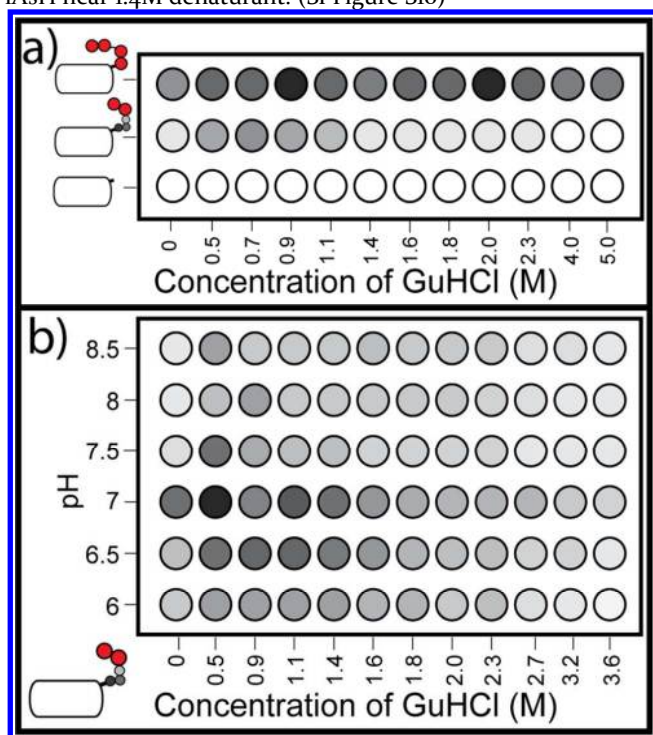
The bipartite designs, **DpsGGGCC** and **DpsPAGCC**, displayed strong, oligomerization dependent binding of FAsH with the latter generating ~50% of control. It should be noted that the bipartite mutants halve the number of cysteine residues and therefore would be expected to have fewer available binding sites per protein, but, with that said, we have not yet rigorously established saturation stoichiometry. These data again demonstrate that at least one binding site design was successful and that our “quick and dirty” lysate-based screen generates accurate results indicating the potential for further optimization and expanding the throughput.

The steady state fluorescence experiments, coupled with denaturation, strongly demonstrate that we have successfully

designed FAsH binding sites that are oligomerization dependent. While those experiments, combined with the geometrical placement of the binding sites, suggested that this dependence is specific to the cage state over the dimer state, it needed to be confirmed directly. We have previously shown that some ferritins can exist in solution as mixtures of cage and dimer through the use of SEC.<sup>28</sup> Therefore we employed SEC, combining protein absorbance (280 nm) with FAsH fluorescence (535 nm). If the binding sites were successfully designed to appear during the formation of specific protein-protein interactions, fluorescent peaks should correspond to only certain oligomerization forms. (Figure 3) The negative control, **Dps**, which is a wild type protein we have worked with extensively, generates a single peak from the cage and no detectable dimer when monitored at 280 nm. (Figure 3b) Also as expected, no peak is observed in the FAsH channel. For the positive control, **DpsCCPGCC**, a cage and a small dimer peak at 280 nm are observed, and the fluorescence elutes with both, again as expected, indicating no preference for oligomerization state. (Figure 3a) The proteins with the bipartite binding sites, **DpsPAGCC** and **DpsGGGCC**, both show a cage and a dimer peak at 280 nm however, the fluorescence only elutes with the cage, suggesting that the designed binding sites are forming only upon cage formation and that this assay can distinguish between

oligomerization states. (Figures 3c and 3d) It should be noted that **DpsPAGCC** did aggregate slightly. The fact that the aggregate peak also had no fluorescence further demonstrates the robustness of this approach. As a further test of specificity, the same technique was used to assess clarified lysates for **DpsPAGCC** and **DpsGGGCC**. (Figures 3e and 3f) Again, fluorescence elutes at a volume consistent with the cage state. Taken together these data strongly suggest that not only are the designed binding sites forming upon oligomerization, they form only upon cage formation.

An additional goal of this technique is not only to develop a direct screen for specific oligomerization states of protein cages, but also to expand the throughput of this characterization. Therefore it was modified to a 96-well plate format. As a proof of principle, but also as a means to optimize our work with these proteins (see above), we sought to discover conditions of highest stability. First we expanded our denaturation experiment described in Figure 1 to screen a variety of denaturant concentrations. (Figure 4a) As expected, and consistent with our previous data, the positive control, **DpsCCPGCC**, generated high signal at all conditions, demonstrating that FAsH binding is not oligomerization dependent. The negative control, **Dps**, generates no signal across all conditions. However, a protein with the designed bipartite binding site, **DpsPAGCC**, loses its ability to bind FAsH near 1.4M denaturant. (SI Figure S10)



**Figure 4** – Medium throughput screen of buffer conditions favoring cage oligomerization state. Fluorescence intensity is greyscaled so that black is most intense and white is least. (a) Normalized fluorescence intensity of **DpsCCPGCC**, **DpsPAGCC** and **Dps** as a function of GuHCl concentration at pH 7.8. Values are the average of at least six replicates. S.D. is included in Supporting Information (SI Figure S10) (b) Normalized fluorescence intensity for **DpsPAGCC** as a function of both pH and GuHCl. Values are averages of three

replicates. S.D. is included in Supporting Information (SI Figure S9)

Further expanding our enquiry into ideal conditions for **Dps** stability, we screened both denaturant and pH concurrently. (Figure 4b and SI Figure S9) It should be noted that, consistent with most of the literature,<sup>35,37</sup> we usually preform our experiments with **Dps** at pH 7.8. Thus, it was somewhat surprising to discover that the protein is most stable below pH 7.0. These experiments would have been less possible in lower throughput screens as they would have taken longer and expended a large amount of reagents. This not only emphasizes the strength of this approach in that its throughput can be expanded, but this increased throughput can lead to useful experiments (with potentially unexpected results) that may not have been run due to the exigent nature of traditional techniques.

## Conclusion

Protein cages have the potential for applications in fields as wide spanning as drug delivery, catalysis, and nanomaterials. Moreover, they can act as model systems to study biologically ubiquitous protein-protein interactions, self-assembly and quaternary structure all of which are at the cutting edge of pharmaceutical interest. However, for protein cages to reach their full potential, tools to assess directly their solution assembly properties in higher throughput are necessary. By engineering FAsH binding sites and protein-protein interfaces that only form in the cage oligomerization state and not in the presumed dimer intermediate, we designed a system to directly detect assembly in cellular lysates of the miniferritin *E.coli* **Dps**. We have shown that indeed this system is oligomerization dependent and is specific for the cage and not the dimer. In addition, we demonstrated the scalability of this system by performing a two dimensional, medium throughput screen to determine conditions that favor the cage state. We are currently expanding this technique to other ferritins in the form of protein libraries screened in whole bacteria for the purpose of discovering “switchable” protein cages and those with bespoke properties for specific applications like nanoparticle formation while optimizing the designs based on ideal binding sites.<sup>38</sup> In addition, we are intending to use this technique to target specific oligomerization states in order to monitor their formation along the self-assembly pathway. We believe that this strategy can be easily ported to other protein cages and self-assembling protein systems in order to optimize their properties and to understand their formation.

## Methods and Materials

**FAsH Binding in Lysates.** The pET-22b expression vectors containing **Dps** design variants (SI Methods and materials S1 for cloning information) were transformed into Rosetta *E.coli* cells (Novagen) and plated on LB agar plates (50  $\mu$ l/ml of carbenicillin and 34  $\mu$ l/ml chloramphenicol). Selected colonies then were grown in LB (3 ml with 50  $\mu$ l/ml of carbenicillin, 37  $^{\circ}$ C, overnight) as a pre-culture which was then added to LB (100 ml) and grown (37  $^{\circ}$ C) until an O.D.<sub>600</sub> of 0.6. Protein expression was then induced by the addition of IPTG (50  $\mu$ l of a 1 M stock) and the culture was further incubated (3 h, 30  $^{\circ}$ C). The cells were isolated by centrifugation (4,000 rpm, 15 min at 4  $^{\circ}$ C). The cell pellet was resuspended in

FLAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) and sonicated (Misonix, Ultrasonic cell disruptor, pulsed 5 s on 5 s off for 5 min). The protein solution was clarified by centrifugation (15,000 rpm, 45 min, 4 °C) and then filtered (Sartorius, 0.2 µm).

The protein concentration was determined (BCA, Novagen) and the cell lysate was diluted to 1 mg/ml with FLAsH buffer. To insure that all the 1 mg/ml samples had similar amounts of the desired protein, each was analyzed by SDS PAGE (SI Figure S3). Each fluorescence experiment contained protein lysate (200 µl, 0.1 mg/ml) in FLAsH buffer, TCEP (Sigma, final concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) were added and the solution was incubated (room temperature, 2 h) followed by the addition of FLAsH-EDT<sub>2</sub> (Invitrogen, final concentration of 0.1 µM) followed by a further incubation (room temperature, 2 h) in the dark. Each lysate sample was tested in a black Corning 96 well plate in a PerkinElmer Envision 2101 multilabel plate reader, with each design being expressed three times and each expression being tested in 6 different wells and re read twice (Ex filter 485 nm bandwidth 14 nm, Em filter 535 nm bandwidth 25 nm). For the denatured experiments, the above was repeated but with the extra addition of 6 M guanidine and incubated for 2 hours prior to the addition of TCEP, EDT and 2-ME, with protein concentration remaining the same as in previous un-denatured experiments.

**FLAsH Binding with Purified Proteins.** The pET-32b vectors containing the Dps variants (SI Methods and materials S2 for cloning information) were transformed into Rosetta *E.coli* cells (Novagen) and plated on LB agar plates (50 µl/ml of carbenicillin and 34 µl/ml of chloramphenicol). Selected colonies were then grown in LB (5 ml, 37 °C, overnight) as a pre-culture which was added to LB (500 ml) and grown (37 °C) until an O.D<sub>600</sub> of 0.6. Protein expression was then induced by the addition of IPTG (250 µl of a 1 M stock) and the cultures were further incubated (3 h, 30 °C). The cells were isolated by centrifugation (4,000 rpm, 20 min, 4 °C). The cell pellet was resuspended in lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 40 mM Imidazole, 1 mM EDTA, pH 8). Cellytic (10x, Sigma) was added and the solution was incubated (20 min, on ice) and then sonicated (Misonix, ultrasonic cell disruptor, pulsed 5 s on, 5 s off for 5 min). The protein solution was clarified by centrifugation (15,000 rpm, 45 min at 4 °C) and then filtered (Sartorius, 0.2 µm). The protein was purified via affinity purification (GE, Histrap FF, 5 ml, (wash buffer-40 mM Imidazole, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 7.4), (elution buffer-500 mM Imidazole, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 7.4)). Enterokinase digestion (NEB 2 µg/ml) was performed to cleave off the peptide tag from the protein of interest followed by a second Histrap (GE, Histrap HP, 5 ml, (wash buffer-50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 7.4), (elution buffer-500 mM Imidazole, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 7.4)) to remove the tag from solution. The protein solution was further purified by size exclusion chromatography (GE Hiloal 16/60 Superdex, running buffer-50 mM NaH<sub>2</sub>PO<sub>4</sub>). Each protein was then placed into FLAsH buffer via ultrafiltration (Millipore).

Each purified protein was tested by incubating the protein (200 µl of 0.1 mg/ml) in FLAsH buffer with TCEP (Sigma, final

concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) and left to incubate (2 h, room temperature) followed by the addition of FLAsH-EDT<sub>2</sub> (Invitrogen, final concentration of 0.1 µM, 2 h, room temperature). Each sample was prepared directly into a black Corning 96 well plate which was tested in a PerkinElmer Envision 2101 multilabel plate reader (Ex filter 485 nm bandwidth 14 nm, Em filter 535 nm bandwidth 25 nm). Each pure protein was tested six times and reread three times. For the denaturant experiment, each protein was tested as described above but with the addition of 6 M guanidine, while keeping the protein concentration the same and was incubated for 2 hours prior to the addition of TCEP, 2-ME, and EDT.

**Analytical Size Exclusion Chromatography (SEC).** The samples (0.5 ml of 1 mg/ml) in FLAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) were injected onto the column (GE Superdex 200 10/300 GL) at 0.5 ml/min with each protein repeated 3 times. The column was calibrated using six proteins as standards (GE Biosystems Calibration Kit). (See SI Figure S7.)

For fluorescent monitoring, samples were prepared by adding to pure protein (0.5 ml of 1 mg/ml) in FLAsH buffer, TCEP (Sigma, final concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) and incubated (2 h, room temperature). FLAsH-EDT<sub>2</sub> dye was added and incubated in the dark (Invitrogen, final concentration of 0.4 µM, 2 h, room temperature). This sample was desalted (GE, HiTrap 5 ml Desalting column) before injection onto a SEC column (GE Superdex 200 10/300 GL). Samples were taken every 200 µl during the elution and placed into a black Corning 96 well plate. This plate was tested in a PerkinElmer Envision 2101 multilabel plate reader (Ex filter 485 nm bandwidth 14 nm, Em filter 535 nm bandwidth 25 nm). Each purified protein was tested three times with each plate reread three times.

**Transmission Electron Microscopy (TEM).** TEM was performed on a FEI, Tecnai G<sup>2</sup> 20, electron microscope set at 200 KeV. Proteins were immobilized on Formvar/carbon coated 3.05 mm copper grids (TAAB) and negatively stained with 1 % Uranyl acetate.<sup>7</sup> Micrographs were analyzed using ImageJ.<sup>41</sup> (See SI Figure S8.)

**Circular Dichroism Spectroscopy (CD).** The purified proteins were used in FLAsH buffer (100 mM Tris, 100 mM NaCl, 1 mM EDTA, pH 7.8). This experiment was performed on a Applied Photophysics LTD Chirscan spectrometer in a range of 200 nm to 260 nm with a protein concentration of 0.2 mg/ml with a path length of 0.5 mm. Thermal melts were performed on all purified proteins (0.2 mg/ml in FLAsH buffer) in a range of 4 to 85 degrees Celsius.<sup>42, 43</sup> (See SI Figure S4, S5 and Table S4.)

**Electrospray Mass Spectrometry.** Purified proteins were desalted (GE, Hitrap 5ml Desalting column) and analyzed on a Bruker MaXis mass spectrometer after the addition of 1 % Formic acid. (See SI Table S5.)

**pH Vs GuHCl measurement.** Each purified protein was analyzed by incubating the protein (200 µl of 0.1 mg/ml) in a buffer at the relevant pH (pH 6-7, 100 mM Citrate-phosphate. pH 7.5-8.5 100 mM Tris.HCl) and with the correct concentra-



tion of GuHCl, with TCEP (Sigma, final concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) and incubated (2 h, room temperature) followed by the addition of FIAH-EDT<sub>2</sub> (Invitrogen, final concentration of 0.1 μM, 2 h, room temperature). Each sample was prepared directly in a black Corning 96 well plate which was scanned in a PerkinElmer Envision 2101 multilabel plate reader (Ex filter 485 nm bandwidth 14 nm, Em filter 535 nm bandwidth 25 nm). Each protein was prepared three times separately and each was reread three times.<sup>44</sup>

## ASSOCIATED CONTENT

**Supporting Information.** This includes SDS-PAGE for lysate samples, SDS-PAGE for all purified proteins as well as TEM, CD, SEC, and electrospray mass spectrometry data for their characterization and sequence alignments and primers used in their manufacture. Raw data for the pH vs GuHCl experiment is also found here. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*brendan.ornor@kcl.ac.uk.

### Author Contributions

†These authors contributed equally.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENT

We thank E. Hobart, N. Luedtke, Y. Zhang, M. Ardejani, and M. Pecuh for insightful conversations, Fan Rongli for the plasmid coding for WT Dps in pET-32b, and A. T. Phan for the encouragement to perform the experiment in Figure 3. We also thank KCL's Centre for Ultrastructural Imaging and Centre for Bimolecular Spectroscopy for help in characterization of the proteins, D. Thurston's lab for help and access to instrumentation, and J. McDonnell for help with the ESI. T.A.C was sponsored by SINGA and BSE scholarships at NTU and King's respectively and J.F. was supported by CBC's undergraduate research fund. The research at NTU was supported by an SPMS start-up grant, a Singapore Ministry of Education Academic Research Fund Tier 1 Grant (RG 53/06) and B.P.O's personal salary. At KCL it was supported by start up funds from the School of Biomedical Sciences and a Marie Curie CIG: PCIG13-GA-2013-618538.

## ABBREVIATIONS

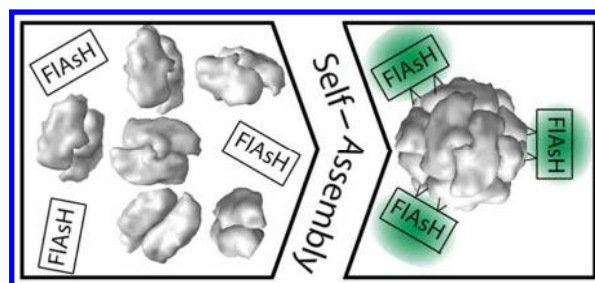
Dps, DNA-binding protein from starved cells; WT Dps, wild type DNA-binding protein from starved cells; FIAH-EDT<sub>2</sub>, 4',5'-bis(1,3,2-dithioarsolan-2-yl)fluorescein-(1,2-ethanedithiol)<sub>2</sub>; GUHCl, Guanidine hydrochloride; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; EDT, 1,2-ethanedithiol; TCEP, tris(2-carboxyethyl)phosphine; 2-ME, 2-mercaptoethanol; SEC, size exclusion chromatography; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; GFP, green fluorescent protein; PCR, polymerase

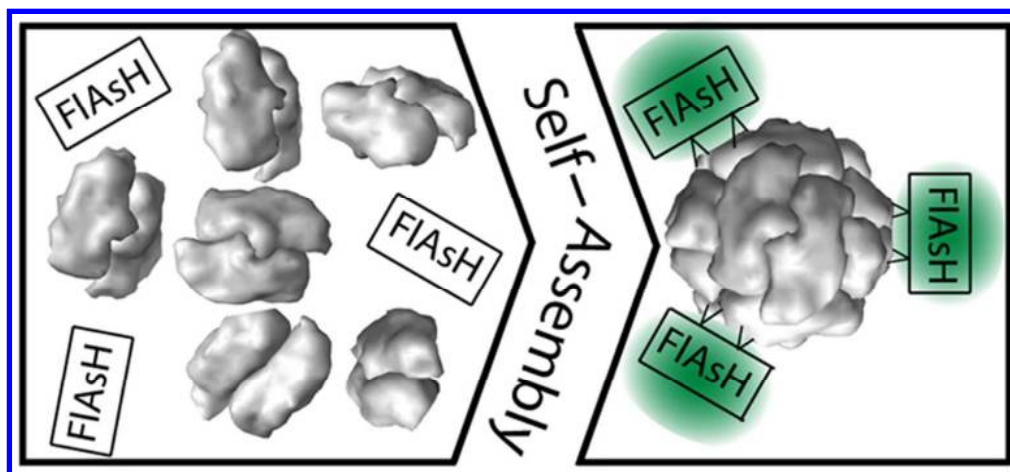
chain reaction; LB, Luria broth; EDTA, Ethylenediaminetetraacetic acid; BCA, bicinchoninic acid; IPTG, Isopropyl β-D-1-thiogalactopyranoside; Ek, Enterokinase; TEM, transmission electron microscopy; CD, circular dichroism; FPR, Fluorescent plate reader; GE, General Electric; Ex, Excitation; Em, Emission; NEB, New England Bioscience.

## REFERENCES

- (1) Grueninger, D.; Treiber, N.; Ziegler, M. O. P.; Koetter, J. W. A.; Schulze, M.-S.; Schulz, G. E. *Science* **2008**, *319*, 206-209.
- (2) Fan, R.; Boyle, A. L.; Cheong, V. V.; Ng, S. L.; Orner, B. P. *Biochemistry* **2009**, *48*, 5623-5630.
- (3) Walter, S.; Buchner, J. *Angew. Chem., Int. Ed.* **2002**, *41*, 1098-1113.
- (4) Aisen, P.; Listowsky, I. *Annu. Rev. Biochem.* **1980**, *49*, 357-393.
- (5) Kis, K.; Volk, R.; Bacher, A. *Biochemistry* **1995**, *34*, 2883-2892.
- (6) Homa, F. L.; Brown, J. C. *Rev. Med. Virol.* **1997**, *7*, 107-122.
- (7) Fan, R.; Chew, S. W.; Cheong, V. V.; Orner, B. P. *Small* **2010**, *6*, 1483-1487.
- (8) Zhang, L.; Swift, J.; Butts, C. A.; Yerubandi, V.; Dmochowski, I. J. *Inorg. Biochem.* **2007**, *101*, 1719-1729.
- (9) Ueno, T.; Suzuki, M.; Goto, T.; Matsumoto, T.; Nagayama, K.; Watanabe, Y. *Angew. Chem., Int. Ed.* **2004**, *43*, 2527-2530.
- (10) Kramer, R. M.; Li, C.; Carter, D. C.; Stone, M. O.; Naik, R. R. *J. Am. Chem. Soc.* **2004**, *126*, 13282-13286.
- (11) Woersdoerfer, B.; Woycechowsky, K. J.; Hilvert, D. *Science* **2011**, *331*, 589-592.
- (12) Keyes, J. D.; Hilton, R. J.; Farrer, J.; Watt, R. K. *J. Nanopart. Res.* **2011**, *13*, 2563-2575.
- (13) Qiu, H. J.; Dong, X. C.; Sana, B.; Peng, T.; Paramelle, D.; Chen, P.; Lim, S. *ACS Appl. Mater. Interfaces* **2013**, *5*, 782-787.
- (14) Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M. *Nat. Struct. Biol.* **1998**, *5*, 294-303.
- (15) pdb: 1F33
- (16) Ardejani, M. S.; Li, N. X.; Orner, B. P. *Biochemistry* **2011**, *50*, 4029-4037.
- (17) Ardejani, M. S.; Chok, X. L.; Foo, C. J.; Orner, B. P. *Chem. Commun.* **2013**, *49*, 3528-3530.
- (18) Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.; Miles, M. J.; Sessions, R. B.; Verkade, P.; Woolfson, D. N. *Science* **2013**, *340*, 595-599.
- (19) Woolfson, D. N. *Adv. Protein Chem.* **2005**, *70*, 79-112.
- (20) Zhang, Y.; Raudah, S.; Teo, H.; Teo, G. W. S.; Fan, R.; Sun, X.; Orner, B. P. *J. Biol. Chem.* **2010**, *285*, 12078-12086.
- (21) Zhang, Y.; Fu, J.; Chee, S. Y.; Ang, E. X. W.; Orner, B. P. *Prot. Sci.* **2011**, *20*, 1907-1917.
- (22) Haile, D. J. *Am. J. Med. Sci.* **1999**, *318*, 230-240.
- (23) Harrison, P. M.; Arosio, P. *Biochim. Biophysica Acta, Bioenerg.* **1996**, *1275*, 161-203.
- (24) Hintze, K. J.; Theil, E. C. *Cell. Mol. Life Sci.* **2006**, *63*, 591-600.
- (25) Leipuvienė, R.; Theil, E. C. *Cell. Mol. Life Sci.* **2007**, *64*, 2945-2955.
- (26) Theil, E. C. *Annu. Rev. Biochem.* **1987**, *56*, 289-315.
- (27) Fan, R.; Boyle, A. L.; Vee, V. C.; See, L. N.; Orner, B. P. *Biochemistry* **2009**, *48*, 5623-5630.
- (28) Zhang, Y.; Orner, B. P. *Int. J. Mol. Sci.* **2011**, *12*, 5406-5421.

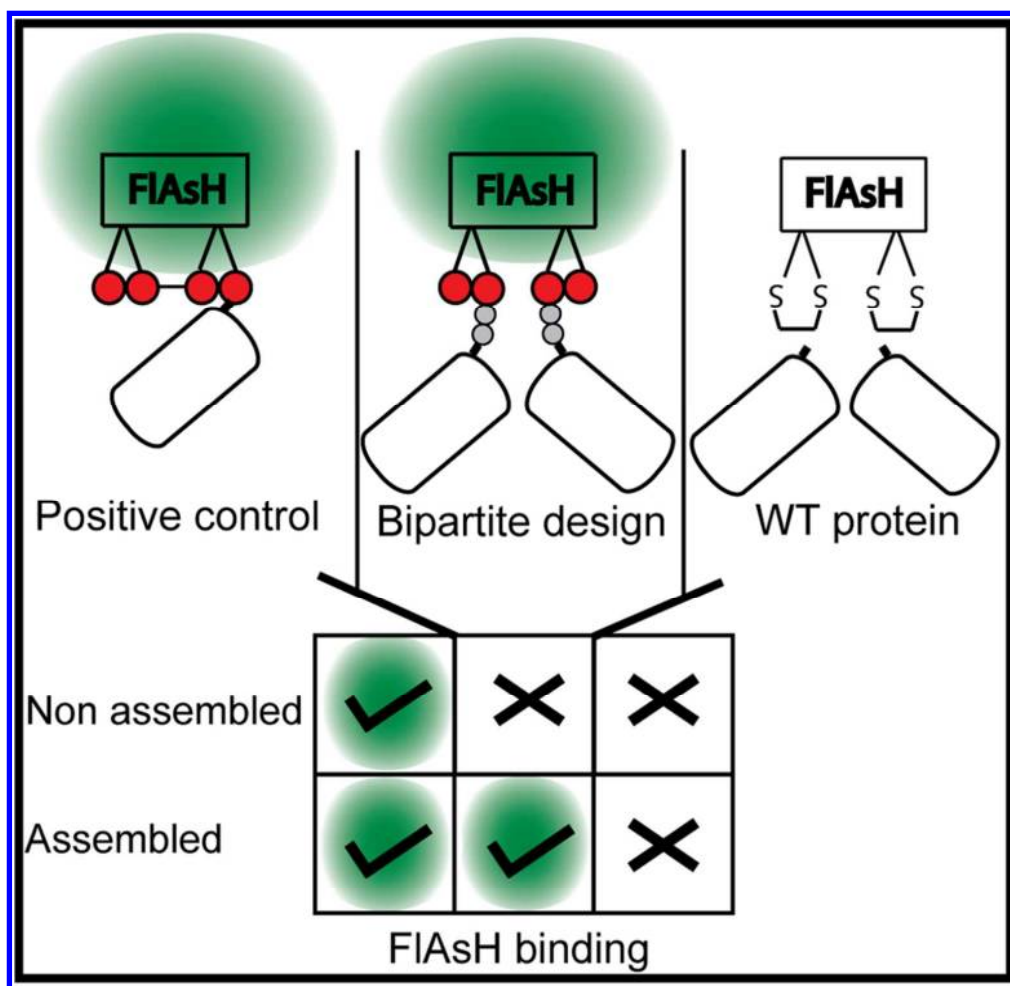
- (29) Griffin, B. A.; Adams, S. R.; Tsien, R. Y. *Science* **1998**, *281*, 269-272.
- (30) Griffin, B. A.; Adams, S. R.; Jones, J.; Tsien, R. Y. *Methods Enzymol.* **2000**, *327*, 565-578.
- (31) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509-544.
- (32) Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y. *J. Am. Chem. Soc.* **2002**, *124*, 6063-6076.
- (33) Ignatova, Z.; Gierasch, L. M. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 523-528.
- (34) Ray-Saha, S.; Schepartz, A. *ChemBioChem* **2010**, *11*, 2089-2091.
- (35) Scheck, R. A.; Schepartz, A. *Acc. Chem. Res.* **2011**, *44*, 654-665.
- (36) Madani, F.; Lind, J.; Damberg, P.; Adams, S. R.; Tsien, R. Y.; Graslund, A. O. *J. Am. Chem. Soc.* **2009**, *131*, 4613-5.
- (37) Luedtke, N. W.; Dexter, R. J.; Fried, D. B.; Schepartz, A. *Nat. Chem. Biol.* **2007**, *3*, 779-784.
- (38) Goodman, J. L.; Fried, D. B.; Schepartz, A. *ChemBioChem* **2009**, *10*, 1644-1647.
- (39) Scheck, R. A.; Lowder, M. A.; Appelbaum, J. S.; Schepartz, A. *ACS Chem. Biol.* **2012**, *7*, 1367-1376.
- (40) Enninga, J.; Mounier, J.; Sansonetti, P.; Van Nhieu, G. T. *Nat. Methods* **2005**, *2*, 959-965.
- (41) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. *Nat. Methods* **2012**, *9*, 671-675.
- (42) Kelly, S. M.; Jess, T. J.; Price, N. C. *Biochim. Biophysica Acta, Proteins Proteomics* **2005**, *1751*, 119-139.
- (43) Greenfield, N. J. *Nat. Protoc.* **2006**, *1*, 2876-2890.
- (44) Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266-80.





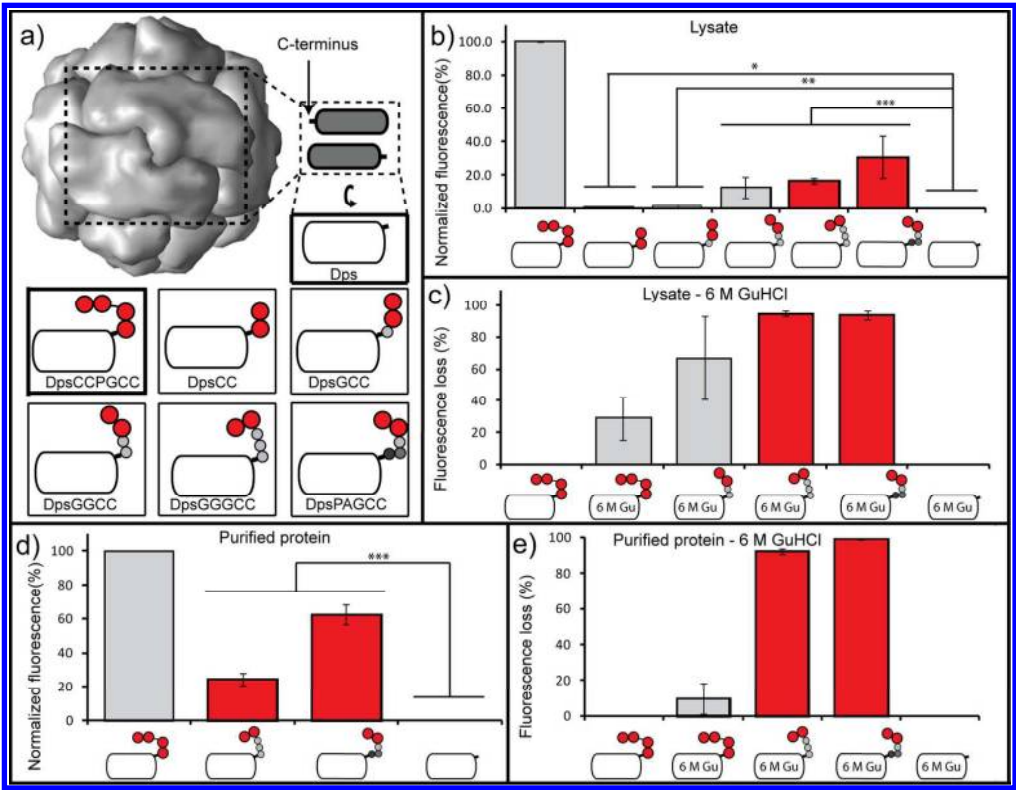
ABSTRACT: Proteins that form cage-like structures have been of much recent cross-disciplinary interest due to their application to bioconjugate and materials chemistry, their biological functions spanning multiple essential cellular processes, and their complex structure, often defined by highly symmetric protein-protein interactions. Thus, establishing the fundamentals of their formation, through detecting and quantifying important protein-protein interactions, could be crucial to understanding essential cellular machinery, and for further development of protein-based technologies. Herein we describe a method to monitor the assembly of protein cages by detecting specific, oligomerization state dependent, protein-protein interactions. Our strategy relies on engineering protein monomers to include cysteine pairs that are presented proximally if the cage state assembles. These assembled pairs of cysteines act as binding sites for the fluorescent reagent FIAsH which, once bound, provides a readout for successful oligomerization. As a proof of principle, we applied this technique to the iron storage protein, DNA-binding protein from starved cells (Dps) from *E. coli*. Several linker lengths and conformations for the presentation of the cysteine pairs were screened to optimize the engineered binding sites. We confirmed that our designs were successful in both lysates and with purified proteins, and that FIAsH binding was dependent upon cage assembly. Following successful characterization of the assay, its throughput was expanded. A two-dimensional matrix of pH and denaturing buffer conditions was screened to optimize nanocage stability. We intend to use this method for the high throughput screening of protein cage libraries and of conditions for the generation of inorganic nanoparticles within the cavity of these and other cage proteins.

76x34mm (300 x 300 DPI)



Fluorescence (green) from the reagent, FIAsH, is assembly state dependent only in the bipartite design because of the need to have two sets of cysteine pairs (red) to form a binding site. The positive control provides a binding site that is neither folding, nor oligomerization dependent whereas the negative control provides no FIAsH binding. This is the basis of our strategy to directly monitor protein cage formation.

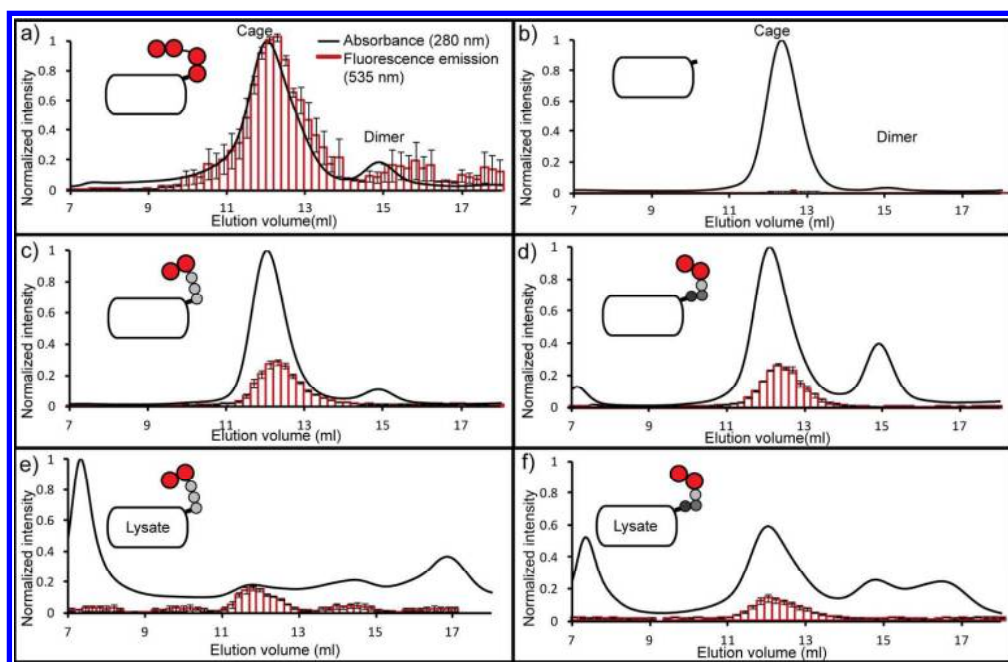
84x81mm (300 x 300 DPI)



(a) Dps crystal structure (PDB: 1DPS) emphasizing a dimer subunit and divergent presentation of the C-termini along with schematics representing proteins used in this study. The negative control (Dps) has no appended C-terminal cysteines whereas the monomer of the positive control (DpsCCPGCC) displays a full binding site. The other proteins present a pair of cysteines with a variety of flexible and constrained linkers. Cysteine (red), glycine (light grey), alanine (dark grey), proline (black) (b) FIAsh fluorescence, normalized to controls, for the proteins overexpressed as cell lysates diluted to the same total protein concentration (0.1 mg/ml). (c) Percent fluorescence loss for the proteins in lysates in denaturing conditions. (d) FIAsh fluorescence, normalized to controls, for the proteins overexpressed and purified, diluted to the same protein concentration (0.1 mg/ml) (e) Percent fluorescence loss for the purified proteins in denaturing conditions. The lysate and purified protein data are from six and eight replicates respectively. Error bars are S.D. \* Two-tailed P-values = 0.4804. \*\* Two-tailed P-values = 0.020050. \*\*\* Two-tailed P-values = 0.0001.

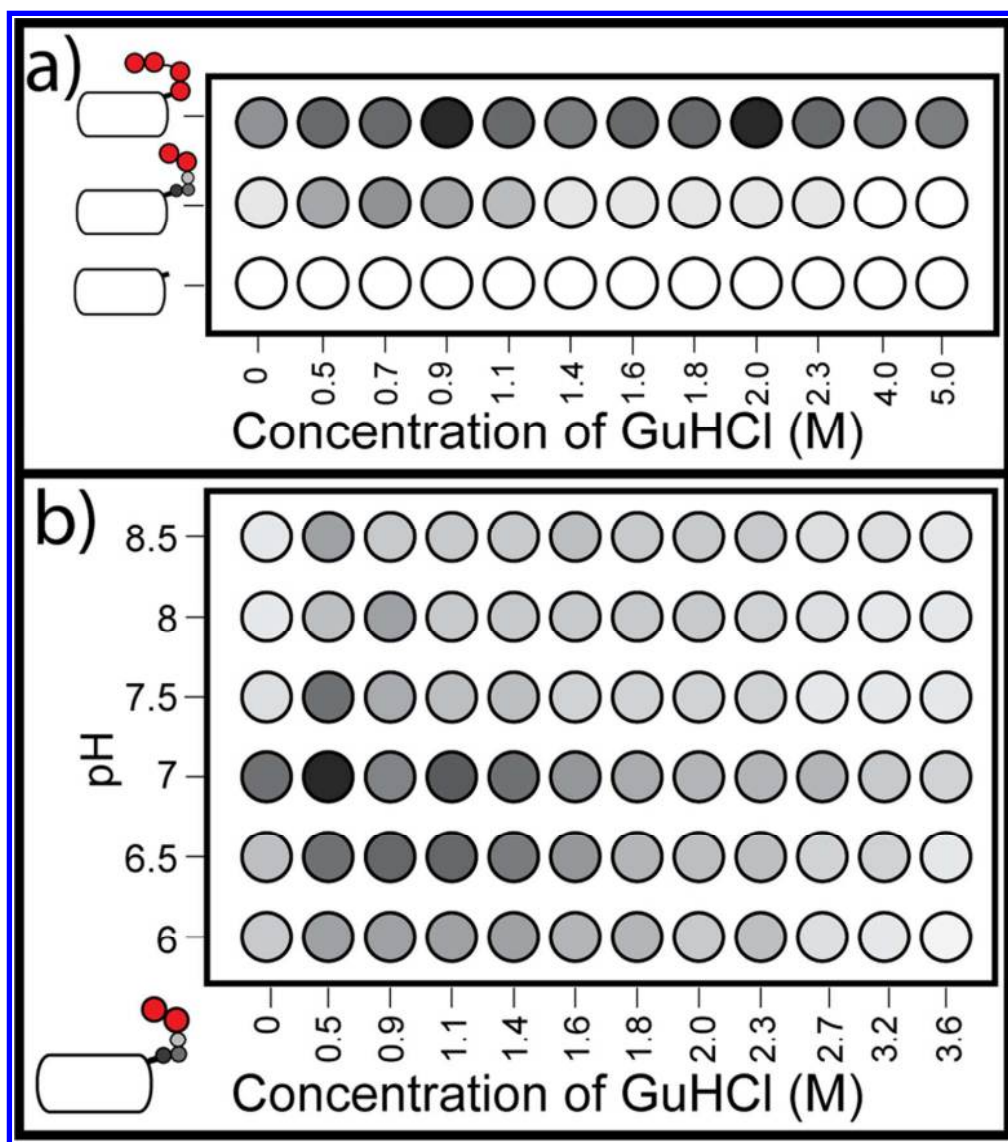
177x137mm (300 x 300 DPI)





Size exclusion chromatograms monitored at absorbance of 280 nm (protein) and fluorescence at 535 nm (bound FIAsh) to determine if the assay is selective for the cage oligomerization state. (a-d) Purified proteins (0.5 ml injection of 1 mg/ml). (e, f) Lysate solutions with overexpressed designed protein (0.5 ml injection of 1 mg/ml total protein). All absorbance and fluorescence signals are normalized to the positive control. Results for both the UV and fluorescence traces are averages of three runs.

177x114mm (300 x 300 DPI)



Medium throughput screen of buffer conditions favoring cage oligomerization state. Fluorescence intensity is greyscaled so that black is most intense and white is least. (a) Normalized fluorescence intensity of DpsCCPGCC, DpsPAGCC and Dps as a function of GuHCl concentration at pH 7.8. Values are the average of at least six replicates. S.D. is included in Supporting Information (SI Figure S9) (b) Normalized fluorescence intensity for DpsPAGCC as a function of both pH and GuHCl. Values are averages of three replicates. S.D. is included in Supporting Information (SI Figure S8)

84x94mm (300 x 300 DPI)